Appl Biochem Biotechnol (2010) 160:1473–1483
DOI 10.1007/s12010-009-8584-1

Exploring the Role of C–H….π Interactions on the Structural Stability of Single Chain "All-Alpha" Proteins

V. Shanthi · K. Ramanathan · Rao Sethumadhavan

Received: 16 January 2009 / Accepted: 18 February 2009 /
Published online: 6 March 2009
© Humana Press 2009

Abstract C–H….π interactions are known to be important contributors to protein stability. In this study, we have analyzed the influence of C–H….π interactions in single chain "all-alpha" proteins. In the data set, a total of 181 C–H….π interactions were observed. The most prominent representatives are the interactions between aromatic C–H donor groups and aromatic π acceptors. Eighty-one percent of the C–H….π interactions between side chain to side chain and remaining19% of the C–H….π interactions were observed between side-chain to side-chain five-member aromatic ring. The donor atom contribution to C–H….π interactions was mainly from Phe, Tyr, and Trp residues. The acceptor atom contribution to C–H….π interactions was mainly from Phe, Tyr, Trp, and His. The highest percentage of C–H….π interactions were observed form Phe residue. The secondary structure preference analysis of all C–H….π interacting residues showed that Phe, Tyr, Trp, and His preferred to be in helix. Long-range C–H….π interactions are the predominant type of interactions in single chain all-alpha proteins data set. All the C–H….π interactions forming residues in the data set preferred to be in the buried region. Seventy-three percent of the donor residues and 65% of the acceptor residues are highly conserved.

Keywords C–H….π interactions · Secondary structure · Interactions range · Solvent accessibility · Conservation

Introduction

The C–H….π interaction is becoming more and more emphasized in modern chemistry, especially in the fields of biochemistry and biophysical chemistry, and it is considered as an important interaction in the stability of three dimensional protein structure. In proteins, C–H….π interactions occur between the C atom of main- or

V. Shanthi · K. Ramanathan · R. Sethumadhavan (✉)
School of Biotechnology, Chemical and Biomedical Engineering, Bioinformatics Division,
Vellore Institute of Technology, Vellore 632014 Tamil Nadu, India
e-mail: rsethumadhavan@vit.ac.in

side-chain amino acid residue and the aromatic side chains of phenylalanine (F), tyrosine (Y), tryptophan (W), and histidine (H). The exothermic dissolution of benzene and similar compounds (π -electron system/proton acceptor) in chloroform (C–H group/proton acceptor) was perhaps the origin of an interaction, now known as C–H….π interactions [1]. In 1957, Reeves and Schneider showed by NMR that this interaction was a type of H bond [2]. Since then, C–H….π interactions have been described in a vast number of small molecule systems from simple olefinic and aromatic compounds to complicated clathrates and inclusion complexes. In 1998, Nishio et al. published excellent treatise of these observations [3]. In this way, C–H….π interactions are gradually gaining a lot of importance. They are a kind of weak hydrogen bonds [1].

The cases in which C–H….π interactions have been described in proteins include the formation of complexes of proteins with special ligands or cofactors such as the heme group [3], pyridoxal-5′-phosphate [4], nucleotides [5, 6], carbohydrates [7] and bound peptides [8], or special geometric circumstances, for instance between neighboring side-chains around a *cis* peptide bond [9]. The importance of this interaction has also been recognized in the design of serine protease inhibitors [10, 11]. There are also recent reviews [12–14] and monographs [3] where the role of CH/π interactions in the structure of chemical and biological macromolecules are described. These interactions also play an important role in the interaction between protein and lipid membranes [5]. Such a type of continuous upshot prompted us to study the relation between occurrences of C–H….π interactions within the protein to the structural stability. Hence, in this work, an effort has been made to collect the information concerning C–H….π interactions in the structural stability of single chain "all-alpha" proteins. In addition, secondary structure, solvent accessibility, interaction range, and conservation have also been analyzed. For all this study, we have chosen only one chain in the proteins structure. These represent relatively simpler systems in which all the weaker interactions can be studied in the absence of the effects of a complex quaternary structure and the occurrence of redundancy in the data set.

We emphasize that 51 proteins in our data set showed a C–H….π interactions, and hence, we accentuate that this investigation is very significant in the sense that C–H….π interactions in single chain all-alpha proteins do play a major role in structural stability of these proteins. It is noteworthy to mention here that the percentage of C–H….π interactions was higher than the percentage of cation–π interactions in the same set of proteins studied by Martis [15]. Hence, without ambiguity, we can confirm that C–H….π interactions play an important role in the structural stability of single chain all-alpha proteins.

Materials and Methods

Data Set

We have selected a set of 75 non-redundant single chain all-alpha proteins, with sequence identity less than 25%, using the sequence analysis package EMBOSS. EMBOSS is an EBI tool that can be used for pair wise alignment of protein sequences. The co-ordinates of the proteins have been taken from the PDB [16]. The PDB codes of the proteins used for the analysis are shown in Table 1. The single chain all-alpha proteins have been selected from the following fivefolds as classified by the SCOP [17]: (1) cytochrome c (a.3), (2) DNA/RNA binding 3-helical bundle (a.4), (3) four helical up and down bundle (a.24), (4) fold/EF hand like (a.39), and (5) alpha–alpha super helix (a.118).

Table 1 List of PDB codes of single chain "all-alpha" proteins considered for analysis of C–H…. π interactions.

a.3 Fold	a.4 Fold	a.24 Fold	a.39 Fold	a.118 Fold
1A56	1AOY	1A7D	1CDP	1B89
1C52	1C20	1CGN	1E14	1EYH
1C2N	1D5V	1CPQ	1IG5	1HF8
1CC5	1D8J	1DOV	1IJ5	1HO8
1CCH	1G2H	1G5Z	1K9P	1HU3
1CCR	1GVD	1GS9	1MHO	1HZ4
1CRY	1GXQ	1KTM	1Q80	1IB2
1CYJ	1IG6	1LPE	1RK9	1KLX
1E8E	1JGS	1NFN	1RRO	1LRV
1F1F	1LEA	1NZE	1SRA	1M8Z
1GDV	1LFB	1O3U	1TOP	1OYZ
1GKS	1MIJ	1SR2	2SAS	1PBV
1LS9	1P4W	1TQG	3PAT	1PAQ
1YCC	2EZI	2A0B	5PAL	1TE4
451C	2HTS	2MHR	5TNC	2BCT

C–H….π Interactions

C–H….π interactions are calculated using the program available for this purpose called HBAT [18]. The C–H….π interactions considered here were between all possible donor C–H groups in protein structures (C^{α}–H, C^{ali}–H, and C^{aro}–H) and between all side-chain π systems (the aromatic rings of Phe, Tyr, Trp, and His). The positions and geometry of donor and acceptor atom with their default parameters are shown in Fig. 1. The donor group is represented as C–H, and the acceptor is the π system. The distances are usually measured from the centroid (M), i.e., center of the π ring. P1 and P2 are distances from C and H, respectively, to M. P3 is the angle between vectors C–H and H–M while P4 is the angle between the CM and MN. Here, N is a normal to the center of the π ring. The geometry is adapted from earlier work of babu [19]. The C–H….π interaction types are represented by a two-letter code in which the first letter indicates the donor atom and the second the acceptor: M, S, and S5 represent the main-chain atom, side-chain atom, and side-chain atom in the five-membered aromatic ring, respectively. We classified the C–H….π interactions into four types of C–H….π interactions, namely, main-chain to side-chain C–

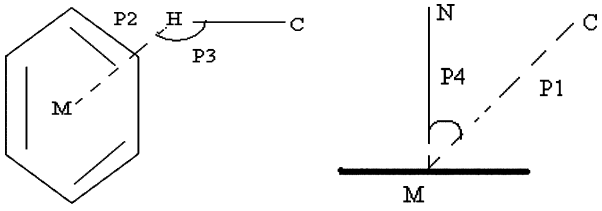

Fig. 1 Parameters for CH/π interactions: P1≤5.00 Å; P2≤4.50 Å; P3≥120°; P4≤30°

H….π interactions (MS–C–H….π), main-chain to side-chain five-member aromatic ring C–H….π interactions (MS5–C–H….π), side-chain to side-chain C–H….π interactions (SS–C–H….π), and side-chain to side-chain five-member aromatic ring C–H….π interactions (SS5–C–H….π) [19].

Secondary Structure and Solvent Accessibility

Secondary structure and solvent accessibility are the two major intermediate steps to understand the structure and function of proteins. Hence, a systematic analysis of each C–H….π interaction forming residue was performed based on their location in different secondary structures of all-alpha proteins and their solvent accessibility. We obtained the information about secondary structures and solvent accessibility of the proteins using the program DSSP [20]. Solvent accessibility was divided into three classes, buried, partially buried, and exposed, indicating, respectively, the least, moderate, and high accessibility of the amino acid residues to the solvent [21, 22].

Sequential Distance

The C–H….π interacting residues coming within a sphere of 8 Å was computed as described earlier [23–25]. For a given residue, the comparison of the surrounding residue is analyzed in terms of the location at the sequence level. The residues that are within a distance of two residues are considered to contribute to short-range interactions, whereas those within a distance of ±3 or ±4 residues contribute to medium range, and those with more than four residues away contribute to long-range interactions [26]. This classification enables us to evaluate the contribution of short-, medium-, and long-range contacts in the formation of C–H….π interactions.

Conservation Score

We computed the conservation score of C–H….π interacting amino acid residues in each protein using the ConSurf server [27]. This server computes the conservation based on the comparison of the sequence of a PDB chain with the proteins deposited in Swiss-Prot [28] and finds the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the *E* value cutoff used in all similarity searches were 1 and 0.001, respectively. All the sequences that are evolutionarily related with each one of the proteins in the data set were used in the subsequent multiple alignments. Based on these protein sequence alignments, the residues are classified into nine categories from highly variable to highly conserved. Residues with a score of 1 are considered highly variable, and residues with a score of 9 are considered highly conserved.

Results and Discussion

C–H….π Interactions

There are four types of C–H….π interactions, namely, main-chain to side-chain interactions (MS–C–H….π), main-chain to side-chain five-member aromatic ring C–H….π interactions (MS5–C–H….π), side-chain to side-chain C–H….π interactions (SS– C–H….π), and side-chain to side-chain five-member aromatic ring C–H….π interactions (SS5–C–H….π) [19].

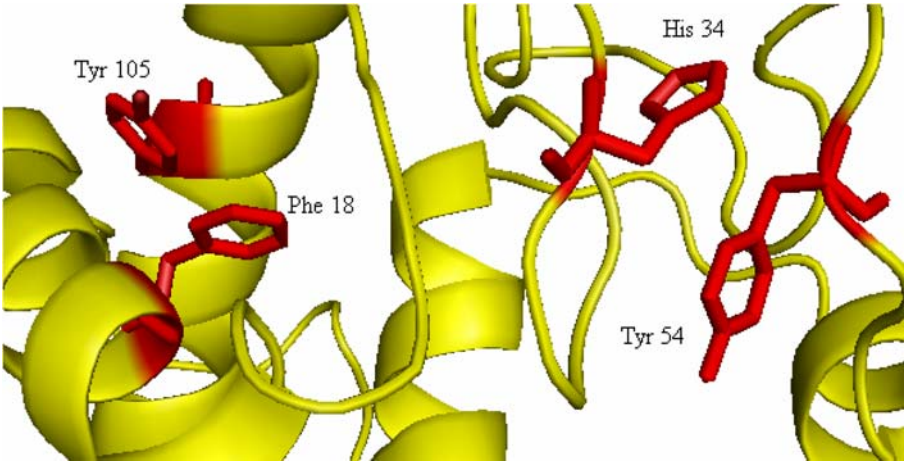

Fig. 2 Pymol view of C–H….π interactions in 1CCR

As a representative picture, the SS–C–H….π interactions and SS5–C–H….π interactions in all-alpha protein PDB ID 1CCR are shown in Fig. 2.

We found that 81% of the interactions were SS–C–H….π, and 19% interactions were SS5–C–H….π. This is shown in Fig. 3. There is no MS–C–H….π and MS5–C–H….π interactions found in the single chain all-alpha proteins data set studied in this work. Among the donor residues, 57% of the interactions were from Phe, 31% of the interactions were from Tyr, and the remaining 12% of the interactions were from Trp residue. In the acceptor π residue, 52% of the interactions were from Phe, 18% of the interactions were from Tyr, 11% of the interactions were from Trp, and 19% of the interactions were from His residue. This is shown in Fig. 4. Hence, Phe residue may be quite important for the stability of single chain all-alpha proteins.

The C–H….π interactions forming residues distance with respect to its position of C atom is shown in Fig. 5. It was found that majority of the 20% of the interactions were found between the residue distances in the range of 4.26 to 4.50 Å. The C–H….π interactions forming residues distance with respect to its position of H atom were shown in Fig. 6. Of the total 181 interactions, the majority of the 26% of the interactions were found between the residue distances in the range of 3.76 to 4.00 Å. Hence, in single chain all-alpha proteins, majority of the C–H….π interactions formed in the distances 4.26 to 4.50 Å

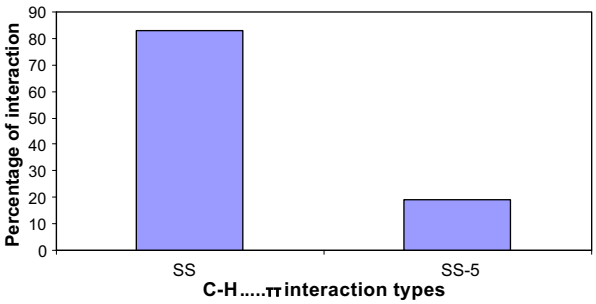

Fig. 3 C–H….π interactions types in single chain all-alpha proteins

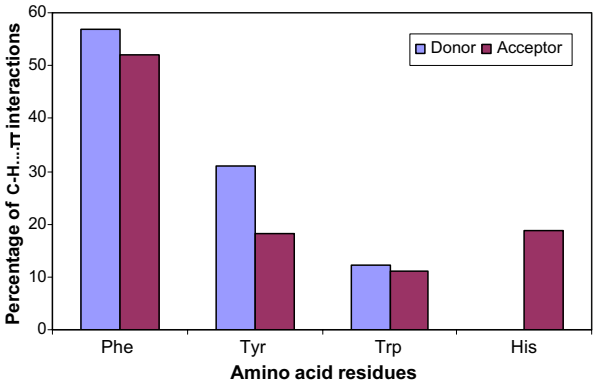

Fig. 4 Contribution amino acid residues in C–H….π interactions

and 3.76 to 4.00 Å from donor C atom and donor H atom, respectively. Among the various types of folds of all-alpha proteins considered for the analysis, proteins belonging to EF hand-like fold had 44% of C–H….π interactions.

Secondary Structure Preferences

The propensity of the amino acid residues to favor a particular conformation has been well documented. Such conformational preference is dependent not only on the amino acid alone but also on the local amino acid sequence. We analyzed the secondary structure preference of each amino acid, which participated in all the different types of C–H….π interactions, namely, SS–C–H….π and SS5–C–H….π interactions. The secondary structure preference of each of the amino acids involved in all the above said types of C–H….π interactions were obtained using DSSP, and the results are depicted in Table 2. It is interesting to note

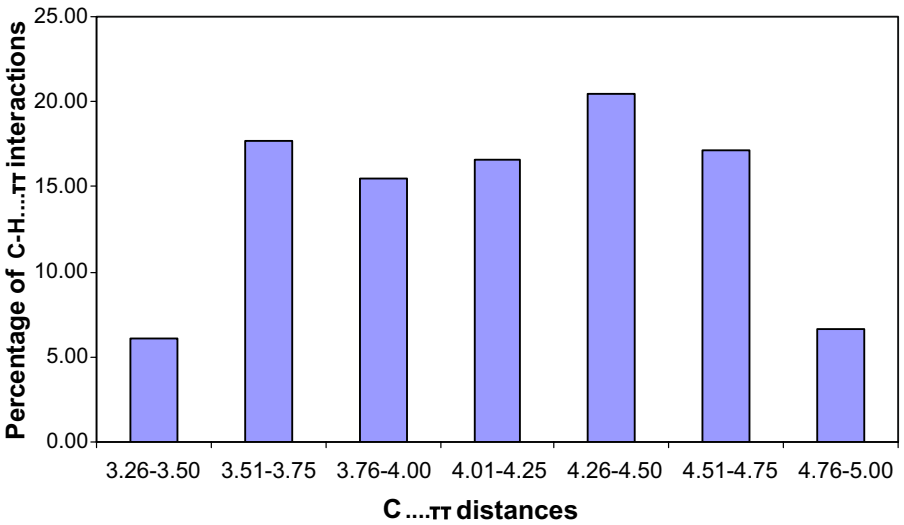

Fig. 5 C….π interacting distances in single chain all-alpha proteins

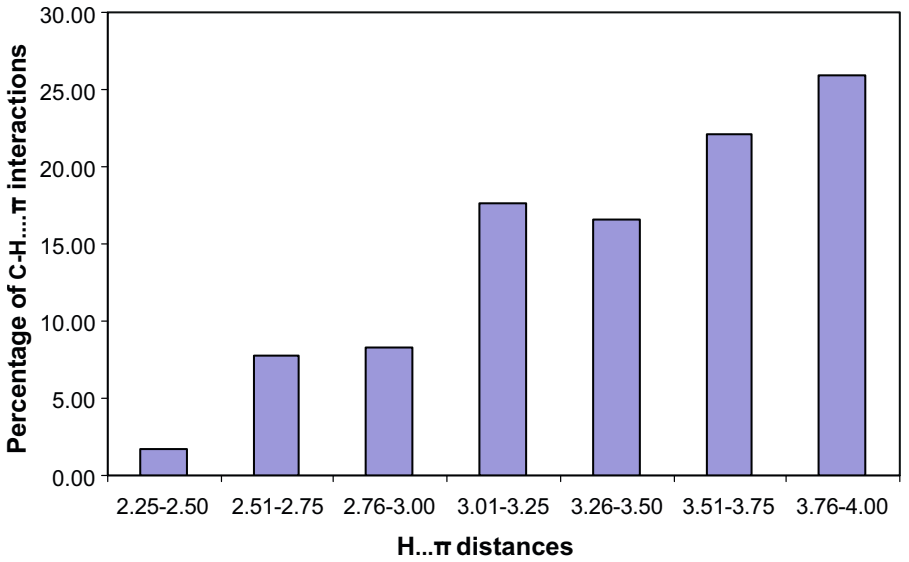

Fig. 6 H….π interacting distances in single chain all-alpha proteins

that all the residues such as Phe, Tyr, Trp, and His were preferred to be in helix. Thus, the single chain all-alpha proteins are therefore confronted with a very large number of helices in their three-dimensional arrangements. Also, we can deduce that residues in helices have a high tendency to form the C–H….π interactions in single chain all-alpha proteins. This data is consistent with the previously suggested information that residues in alpha helices have the tendency to form C–H….π interactions [29].

Solvent Accessibility

The relation between the amino acid residues in C–H….π interactions and solvent accessibility is depicted in Fig. 7. The solvent accessibility of amino acid residues has been categorized as buried, partially buried, and exposed [21, 22]. We found that all the aromatic residues such as Phe, Tyr, Trp, and His residues were preferred to be in buried regions. This observation is quite reasonable in the sense that the aromatic residues are in principle non-polar residues and tend to be buried. According to Weiss et al. [29], C–H….π interactions involving aromatic residues either as donor or as acceptor groups are found mostly in the interior of the protein and tend to be buried in nature. These might be one of the reasons for their nature of solvent accessibility.

Table 2 Frequency of occurrence of C–H….π interaction forming residue in different secondary structures.

Residue	Helix	Coil	Turn
Phe	83.69	13.04	3.27
Tyr	78.26	9.78	11.96
Trp	54.28	40.00	5.71
His	80.64	19.35	Nil

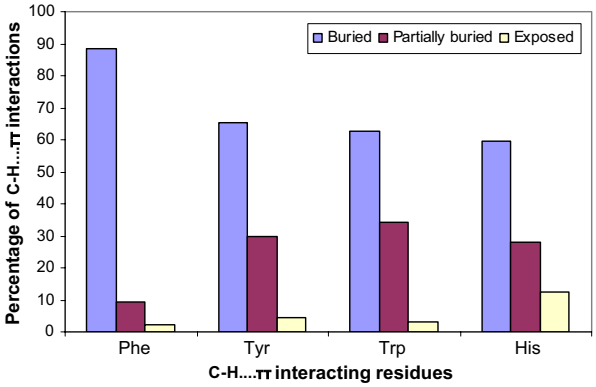

Fig. 7 Solvent accessibility of C–H….π interacting residues in single chain all-alpha proteins

Sequential Separation

The contribution of C–H….π interactions in single chain all-alpha proteins could define either the local or the global stability of the proteins. Therefore, there is a need to evaluate the contribution of inter-residual C–H….π interactions. The residues that are within a distance of two residues are considered to contribute to short-range interactions, whereas those within a distance of ±3 or ±4 residues contribute to medium range, and those with more than four residues away contribute to long-range interactions [26]. This classification enables us to evaluate the contribution of short-, medium-, and long-range contacts in the formation of C–H….π interactions. The sequential distance between residues that contributed to C–H….π interactions were calculated, and results are depicted in Fig. 8; 69%, 18%, and 13% of inter-residue C–H….π interactions were found to be long-, medium-, and short-range interactions, respectively. Long-range C–H….π interactions are the predominant type of interactions in single chain all-alpha proteins. These results indicate that long-range C–H….π interactions contribute significantly to the global conformational stability of single chain all-alpha proteins.

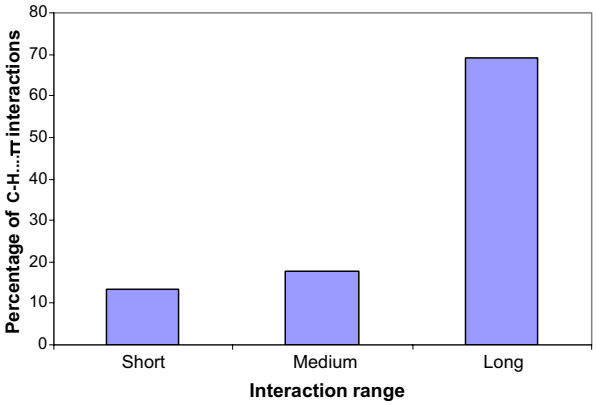

Fig. 8 C–H….π interaction range in single chain all-alpha proteins

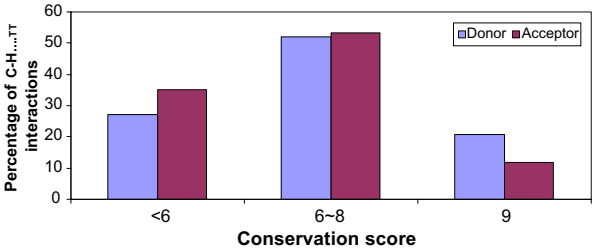

Fig. 9 Conservation score for C–H….π interacting residues in single chain all-alpha proteins

Conservation Score

We used the ConSurf server to compute the conservation score of amino acid residues involved in C–H….π interactions in single chain all-alpha proteins, and the results are shown in Fig. 9. Twenty-one percent of the amino acid residues that contributed donor atoms in C–H….π interactions had the highest conservation score of 9, while 52% of the amino acid residues had a conservation score, in the range of 6–8. Thus, 73% of the donor amino acid residues had a higher conservation score. In the case of amino acid residues that contributed acceptor atoms in C–H….π interactions, 12% of the acceptor amino acid residues had the highest conservation score of 9, while 53% of the amino acid residues had a conservation score, in the range of 6–8. Thus, 65% of the acceptor amino acid residues had a higher conservation score. From these observations, we were able to infer that most of the amino acid residues involved in C–H….π interactions might be conserved in single chain all-alpha proteins. This high conservation of amino acid residues may in some cases be linked to their involvement in C–H….π interactions and to the stability or the function of the protein [29].

Conclusion

We have systematically analyzed the influence of C–H….π interactions to the stability of single chain all-alpha proteins. We found that 68% of the considered single chain all-alpha proteins exhibit C–H….π interactions, and Phe residue plays an important role in forming such interactions. The most prominent representatives are the interactions between aromatic C–H donor groups and aromatic π acceptors. The geometric parameters calculated for these interactions suggest that C–H….π interactions can be classified as weak H bonds and occur mainly in the distances greater than 4.26 and 3.76 Å from the C and H atoms, respectively, in the data set. C–H….π interactions involving aromatic π systems as a donor or acceptor groups are generally found closer to the center of the protein and hence is buried in nature. The secondary structure preference analysis of C–H….π interacting residues showed that all the amino acid residues such as Phe, Tyr, Trp, and His preferred to be in helix confirmation. Thus, we can deduce that residues in helices have high tendency to form the C–H….π interactions in single chain all-alpha proteins. The sequential distance between residues that contributed donor and acceptor atoms confirmed that long-range C–H….π interactions contribute significantly to the global conformational stability of all-alpha proteins. Seventy-three percent of the donor amino acid residues and 65% of the acceptor amino acid residues had a higher conservation score. It might be due to their involvement in

C–H….π interactions and contribute significantly to the stability of single chain all-alpha proteins.

The C–H….π interactions participating groups are not dipolar; this type of interaction persists even in polar protic media such as water, unlike the ordinary hydrogen bond; and Coulombic force tends to be weakened in a manner that is inversely proportional to the dielectric constant of the medium and are obscured in polar solvents. Though weak (around 1 kcal mol^{-1} for a one unit interaction), a unique feature of this force is that many CH groups may participate simultaneously in the interaction with a π base. Total energy of the interaction will increase by organizing CHs and/or π groups into a favorable chemical structure. We conclude that the C–H….π interaction play an important role in the structural stability of antimicrobial peptides.

Acknowledgments The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this work. The authors also thank the reviewers for their suggestions in the improvement of this manuscript.

References

1. Tamres, M. (1952). *Journal of the American Chemical Society, 74*, 3375–3378. doi:10.1021/ja01133a047.
2. Reeves, L. W., & Schneider, W. G. (1957). *Canadian Journal of Chemistry, 35*, 251–261. doi:10.1139/v57-036.
3. Nishio, M., Hirota, M., & Umezawa, Y. (Eds.) (1998). New York: Wiley.
4. Matsui, I., Matsui, E., Sakai, Y., Kikuchi, H., Kawarabayasi, Y., Ura, H., et al. (2000). *The Journal of Biological Chemistry, 275*, 4871–4879. doi:10.1074/jbc.275.7.4871.
5. Chakrabarti, P., & Samanta, U. (1995). *Journal of Molecular Biology, 251*, 9–14. doi:10.1006/jmbi.1995.0411.
6. Umezawa, Y., & Nishio, M. (1998). *Bioorganic & Medicinal Chemistry, 6*, 493–504. doi:10.1016/S0968-0896(98)00002-9.
7. Muraki, M., Harata, K., Sugita, N., & Sato, K. I. (2000). *Biochemistry, 39*, 292–299. doi:10.1021/bi991402q.
8. Umezawa, Y., & Nishio, M. (1998). *Bioorganic & Medicinal Chemistry, 6*, 2507–2515. doi:10.1016/S0968-0896(98)80024-2.
9. Jabs, A., Weiss, M. S., & Hilgenfeld, R. (1999). *Journal of Molecular Biology, 286*, 291–304. doi:10.1006/jmbi.1998.2459.
10. Shimohigashi, Y., Maeda, I., Nose, T., Ikesue, K., Sakamoto, H., Ogawa, Y., et al. (1996). *Journal of the Chemical Society, 1*, 2479–2485.
11. Shimohigashi, Y., Nose, T., Yamauchi, Y., & Maeda, I. (1999). *Biopolymers, 51*, 9–17. doi:10.1002/(SICI)1097-0282(1999)51:1<9::AID-BIP3>3.0.CO;2-5.
12. Steiner, T. (2002). *Angewandte Chemie International Edition in English, 41*, 48–76. doi:10.1002/1521-3773(20020104)41:1<48::AID-ANIE48>3.0.CO;2-U.
13. Steiner, T., & Koellner, G. J. (2001). *Journal of Molecular Biology, 305*, 535–557. doi:10.1006/jmbi.2000.4301.
14. Meyer, E. A., Castellano, R. K., & Diederich, F. (2003). *Angewandte Chemie International Edition in English, 42*, 1210–1250. doi:10.1002/anie.200390319.
15. Martis, R. L., Singh, S. K., Michael Gromiha, M., & Santhosh, C. (2008). *Journal of Theoretical Biology, 250*, 655–662. doi:10.1016/j.jtbi.2007.10.024.
16. Berman, H. M., Westbrook, J. Z., Feng, G., Gilliland, T. N., Bhat, H., Weissig, I. N., et al. (2000). *Nucleic Acids Research, 28*, 235–242. doi:10.1093/nar/28.1.235.
17. Conte, L. L., Ailey, B., Hubbard, T. J. P., Brenner, S. E., Murzin, A. G., & Chothia, C. (2000). *Nucleic Acids Research, 28*, 257–259. doi:10.1093/nar/28.1.257.
18. Tiwari, A., & Panigrahi, S. K. (2007). *In Silico Biology, 7*, 0057.

19. Babu, M. M. (2003). *Nucleic Acids Research, 31*, 3345–3348. doi:10.1093/nar/gkg528.
20. Kabsch, W., & Sander, C. (1983). *Biopolymers, 22*, 2577–2637. doi:10.1002/bip. 360221211.
21. Gilis, D., & Rooman, M. (1996). *Journal of Molecular Biology, 257*, 1112–1126. doi:10.1006/jmbi.1996.0226.
22. Gilis, D., & Rooman, M. (1997). *Journal of Molecular Biology, 272*, 276–290. doi:10.1006/jmbi.1997.1237.
23. Gromiha, M. M., & Selvaraj, S. (1997). *Journal of Biological Physics, 23*, 151–162. doi:10.1023/A:1004981409616.
24. Gromiha, M. M., Santhosh, C., & Ahmed, S. (2004). *International Journal of Biological Macromolecules, 34*, 203–211. doi:10.1016/j.ijbiomac.2004.04.003.
25. Selvaraj, S., & Gromiha, M. M. (2003). *Biophysical Journal, 84*, 1919–1925. doi:10.1016/S0006-3495(03)75000-0.
26. Gromiha, M. M., & Selvaraj, S. (2004). *Progress in Biophysics and Molecular Biology, 86*, 235–277. doi:10.1016/j.pbiomolbio.2003.09.003.
27. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor, D., Martz, E., et al. (2003). *Bioinformatics (Oxford, England), 19*, 163–164. doi:10.1093/bioinformatics/19.1.163.
28. Boeckman, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). *Nucleic Acids Research, 31*, 365–370. doi:10.1093/nar/gkg095.
29. Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J., & Hilgenfeld, R. (2001). *Journal of Molecular Biology, 307*, 357–377. doi:10.1006/jmbi.2000.4473.